

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
13 September 2001 (13.09.2001)

PCT

(10) International Publication Number  
**WO 01/67671 A2**

(51) International Patent Classification<sup>7</sup>: **H04L 12/00**

(21) International Application Number: PCT/IB01/00172

(22) International Filing Date: 12 February 2001 (12.02.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
09/518,875 6 March 2000 (06.03.2000) US

(71) Applicants and

(72) Inventors: **MEYER, Thomas, W.** [US/US]; 315 South Coast Highway, 101, Encinitas, CA 92024 (US). **MEYER, Josslyn, Motha** [CA/US]; 315 South Coast Highway, 101, Encinitas, CA 92024 (US).

(74) Agent: **RINES, Robert, H.**; MacLeod Allsop, Bledington Grounds, Bledington, Gloucestershire OX7 6XL (GB).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— *without international search report and to be republished upon receipt of that report*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: DATA EMBEDDING IN DIGITAL TELEPHONE SIGNALS

(57) **Abstract:** In a cellular telephone system, a novel technique for supplementing the transmission of cell phone voice signals with supplemental advertising, entertainment, e-commerce and service information and the like to be presentable at the user phone handset, involving embedding such supplemental digital data in the digital phone signal without affecting backwards compatibility of the digital phone signal, through transforming the digital voice phone signal into encoded sets of frequency-domain or other transform coefficient representations of said signal, and selecting predetermined coefficient portions that are to contain bits of the supplemental data and embedding such bits at the selected portions while compressing the signal to transmit a compressed digital voice signal containing the supplemental data embedded therein, thus enabling user decoding to extract the supplemental data while receiving the transmitted voice signal.



**WO 01/67671 A2**

## DATA EMBEDDING IN DIGITAL TELEPHONE SIGNALS

Field of Invention

The field of perhaps the most important application of the present invention resides in improved techniques for embedding in digital telephone signals and the like, and in particular cellular phone systems, data supplemental to the voice signal (for example, targeted advertising images, music or other entertainment content, market-localized ad's, interactive e-commerce applications, games, weather and other services). Such embedding is preferably effected at a point where the audio voice signal is being converted from an uncompressed representation to a highly compressed digital representation as part of the coding and compression process before it is transmitted, such as at a user's phone handset as for extraction at a central point, or at a central digitizing point as for extraction at the handset. The invention enables such extraction of the embedded data from the digital voice signal at any point in the process without affecting the digital signal in any way.

Where the supplemental data is intended to be received by a user's phone, for example, it can be extracted into an appropriate format and displayed, executed, stored, or otherwise handled by and/or at the phone; and where the supplemental data is intended to be received at another point in the system, it may be extracted into an appropriate format and acted upon in a manner depending on the semantics of the desired system.

The technique of the invention, furthermore, is also useful to embed such supplemental data at any intermediate point of the system, and including even embedding

where the digital signal has already been compressed; though, in such event, with somewhat less transparency, efficiency and bit rate.

In all applications, however, the invention preferably uses the fundamental techniques disclosed in my earlier joint copending U.S. application Serial Number 09/389,941, filed Sept. 3, 1999, (PCT application No. PCT/IB00,00227), and entitled "Process, System, And Apparatus For Embedding Data In Compressed Audio, Image, Video And Other Media Files And The Like".

### Background

Prior to the invention of my said copending patent application, as explained therein, data has heretofore often been embedded in analog representations of media information and formats. This has been extensively used, for example, in television and radio applications as for the transmission of supplemental data, such as text; but the techniques used are not generally capable of transmitting high bit rates of digital data.

Watermarking data has also been embedded so as to be robust to degradation and manipulation of the media. Typical watermarking techniques rely on gross characteristics of the signal being preserved through common types of transformations applied to a media file. These techniques are again limited to fairly low bit rates. Good bit rates on audio watermarking techniques are, indeed, only around a couple of dozen bits of data encoded per second.

While data has been embedded in the low-bit of the single-domain of digital media enabling use of high bit rates, such data is either uncompressed, or capable of only relatively low compression rates. Many modern compressed file formats, moreover, do not use such signal-domain representations and are thus unsuited to the use of this

technique. Additionally, this technique tends to introduce audible noise when used to encode data in sound files.

Among prior patents illustrative of such and related techniques and uses are U.S. Patents Nos. 4,379, 947 (dealing with the transmitting of data simultaneously with audio); 5,185, 800 (using bit allocation for transformed digital audio broadcasting signals with adaptive quantization based on psychoauditive criteria ); 5,687,236 (steganographic techniques); 5,710, 834 (code signals conveyed through graphic images); 5,832,119 (controlling systems by control signals embedded in empirical data); 5,850,481 (embedded documents, but not for arbitrary data or computer code); 5,889,868 (digital watermarks in digital data); and 5,893, 067 (echo data hiding in audio signals).

Prior publications relating to such techniques include

Bender, W. D. Gruhl, M. Morimoto, and A. Lu, " Techniques for data hiding", *IBM Systems Journal*, Vol. 35, Nos. 3 & 4, 1996, p. 313-336;

MPEG Spec-ISO/IEC 11172, part 1-3, *Information Technology -Coding of moving pictures and associated audio for digital storage media at up to about 1.5Mbit/s* Copyright 1993, ISO/IEC; and

ID3v2 spec: <http://www.id3.org/easy.html> and <http://www.id3.org/id3v2.3.0.html>

A survey of techniques for multimedia data labeling, and particularly for copyright labeling using watermark in the encoding low bit-rate information is presented by Langelaar, G.C. et al. in "Copy Protection For Multimedia Data based on Labeling Techniques"

([http://www-it.et.tudelft.nl/html/research/smash/public/benlx96/benelux\\_cr.html](http://www-it.et.tudelft.nl/html/research/smash/public/benlx96/benelux_cr.html)).

In specific connection with the above-cited "MPEG Spec" and "IDSv2 Spec" reference applications, we have disclosed in co-pending U.S. patent application Serial No. 09/389,942, entitled "Process Of And System For Seamlessly Embedding Executable

Program Code Into Media File Formats Such As MP3 And The Like For Execution By Digital Media Player And Viewing Systems", (PCT application No, PCT/IB00,00227), techniques applying some of the embedding concepts of the present invention, though directed specifically to imbuing one or more of pre-prepared audio, video, still image, 3-D or other generally uncompressed media formats with an extended capability to supplement their pre-prepared presentations with added graphic, interactive and/or e-commerce content presentations at the digital media playback apparatus.

As earlier indicated, however, the technique of my first-named earlier application Serial No. 09/389,941, is more broadly concerned with data embedding in compressed formats, and, indeed, with encoding a frequency representation of the data, typically through a Fourier Transform, Discrete Cosine Transform, or other well-known function. That invention embeds high-rate data in compressed digital representations of the media, including through modifying the low-bits of the coefficients of the frequency representation of the compressed data, thereby enabling additional benefits of fast encoding and decoding, because the coefficients of the compressed media can be directly transformed without a lengthy additional decompression/compression process. The technique also can be used in combination with watermarking, but with the watermark applied before the data encoding process.

The earlier cited Langelaar et al publication, in turn, references and discusses the following additional prior art publications:

J. Zhao, E. Koch: "Embedding Robust Labels into Images for Copyright Protection", Proceedings of the International Congress on Intellectual Property Rights for Specialized Information, Knowledge and New Technologies, Vienna, Austria, August 1995;

E. Koch, J. Zhao: "Towards Robust and Hidden Image Copyright Labeling", Proceedings IEEE Workshop on Nonlinear Signal and Image Processing, Neos Marmaras, June, 1995; and

F. M. Boland, J.J.K O Ruanaidh, C, Dautzenberg: "Watermarking Digital Images for Copyright Protection", Proceedings of the 5th International Conference on Image Processing and its Applications, No. 410, Endinburgh, July, 1995

An additional article by Langelaar also discloses earlier labeling of MPEG compressed video formats:

G.C Langelaar, R. L. Lagendijk, J. Biemond: "Real-time Labeling Methods for MPEG Compressed Video," *18th Symposium on Information Theory in the Benelux*, 15-16 May 1997, Veldhoven, The Netherlands.

These Zhao and Koch, Boland et al and Langelaar et al disclosures, while teaching encoding technique approaches having partial similitude to components of the techniques employed by the present invention, as will now be more fully explained, are not, however, either anticipatory of, or actually adapted for solving the total problems with the desired advantages that are addressed and sought by the present invention.

Considering, first, the approach of Zhao and Koch, above-referenced, they embed a signal in an image by using JPEG-based techniques. ([JPEG] Digital Compression and Coding of Continuous-tone Still Images, Part 1: Requirements and guidelines, ISO/IEC DIS 10918-1. They first encode a signal in the ordering of the size of three coefficients, chosen from the middle frequency range of the coefficients in an 8-block or octet DCT. They divide eight permutations of the ordering relationship among these three coefficients into three groups: one encoding a '1' bit (HML, MHL, and HHL), one encoding a '0' bit (MLH, LMH, and LLH), and a third group encoding "no data" (HLM, LHM, and MMM). They have also extended this technique to the watermarking of video data. While their technique is robust and resilient to modifications, they cannot, however,

encode large quantities of data, since they can only modify blocks where the data is already close to the data being encoded; otherwise, they must modify the coefficients to encode "no data". They must also severely modify the data since they must change large - scale ordering relationships of coefficients. As will later more fully be explained, these are disadvantages overcome by the present invention through its technique of encoding data by changing only a single bit in a coefficient.

As for Boland, Ruanaidh, and Dautzenberg, they use a technique of generating the DCT Walsh Transform, or Wavelet Transform of an image, and then adding one to a selected coefficient to encode a "1" bit, or subtracting one from a selected coefficient to encode a "0" bit. This technique, although at first blush someone superficially similar in one aspect of one component of the present invention, has the very significant limitation, obviated by the present invention, that information can only be extracted by comparing the encoded image with the original image. This means that a watermarked and a non-watermarked copy of any media file must be sent simultaneously for the watermarking to work. This is a rather severe limitation, overcome in the present invention by the novel incorporating of the use of the least-significant bit encoding technique.

Such least-significant bit encoding broadly has, however, been earlier proposed; but not as implemented in the present invention. The Langelaar, Langendijk, and Biemond publication, for example, teaches a technique which encodes data in MPEG video streams by modifying the least significant bit of a variable-length code (VLC) representing DCT coefficients. Langelaar et al's encoding keeps the length of the file constant by allowing the replacement of only those VLC values which can be replaced by another value of the same length and which have a magnitude difference of one. The

encoding simply traverses the file and modifies all suitable VLC values. Drawbacks of their techniques, however, are that suitable VLC values are relatively rare (167 per second in a 1.4 Mbit/sec video file, thus allowing only 167 bits to be encoded in 1.4 million bits of information).

In comparison, the technique of my first-named earlier application Serial No. 09/389,941, as applied for video, for example, removes such limitation and can achieve much higher bit-rates while keeping file-length constant, by allowing a group or set of nearby coefficients to be modified together. This also allows for much higher quantities of information to be stored without perceptual impact because it allows for psycho-perceptual models to determine the choice of coefficients to be modified.

The improved techniques of my earlier invention, indeed, unlike the prior art, allow for the encoding of digital information into an audio, image, or video file at rates several orders of magnitude higher than those previously described in the literature (order of 300 bits per second ). As will later be disclosed, the present invention, indeed, has easily embedded a 3000 bit/second data stream in a 128,000 bit/second audio file.

In the prior art, only relatively short sequences of data have been embedded into the media file, typically encoding simple copyright or ownership information. Our techniques allow for media files to contain entirely new classes of content, such as: entire computer programs, multimedia annotations, or lengthy supplemental communications. As described in said copending applications, computer programs embedded in media files allow for expanded integrated transactional media of all kinds, including merchandising, interactive content, interactive and traditional advertising, polls, e-commerce solicitations such as CD or concert ticket purchases, and fully reactive content such as games and



interactive music videos which, where used with personal computers, react to the user's mouse motions and are synced to the beat of the music. This enables point of purchase sales integrated with the music on such software and hardware platforms as the television, portable devices like the Sony Walkman, the Nintendo Game Boy, and portable MP3 players such as the Rio and Nomad and the like. This creates new business models. For example, instead of a record company trying to stop the copying of its songs, it might instead encourage the free and opened distribution of the music, so that the embedded advertising and e-commerce messages are spread to the largest possible audience and potential customers.

Turning, now, to the present invention, it is directed to applying the above-described techniques of my said earlier patent applications to the specific problems of use with cellular (and other) telephones and the like having very different problems than pre-recorded media, though also useful with pre-recorded voice instead of currently generated real-time voice or other signals to be transmitted over the phones.

#### Objects of Invention

It is accordingly a primary object of the present invention to provide a new and improved process, system and apparatus for embedding supplemental data (such as, for example, advertising images, market-localized ads, interactive computer programs such as e-commerce applications, games, forms, supplemental text or audio content, music or other entertainment content, etc.) on digital cellular (and other) phone signals and without affecting the digital backwards compatibility of the digital phone signal.

A further object is to provide such a novel process in which the embedding involves a single process added at a point where the audio voice signal is converted from an uncompressed representation to a highly compressed digital representation to add the supplemental data to the voice signal as part of the coding and compression process, before it is transmitted.

Still another object is to provide such a novel embedding technique, particularly in a wireless cellular phone system, at the mobile switching center (MSC) or other central point for extraction at the user handset, or at the handset for extraction at the central point.

An additional object is to provide also for the embedding of supplemental data into a digital signal which has already been compressed.

Another object is to provide through the ability to embed supplemental data into the phone signal either at the user's handset for reception at a central station, at the central station for reception at the user's handset, or less efficiently at any intermediate point, the creation of a novel two-way network connection, while the handset is used over a voice-only network.

Other and further objects will be explained hereinafter and are more particularly pointed out in the appended claims.

### Summary

In summary, therefore, from one of its broader aspects, the invention embraces a method of embedding supplemental digital data in a voice digital phone signal without affecting the backwards compatibility of the digital phone signal, that comprises, transforming the digital voice phone signal into encoded sets of frequency-domain or

other transform coefficient representations of said signal; selecting predetermined coefficient portions that are each to contain a bit of the supplemental data; and embedding said bits at the selected portions while compressing the signal to transmit a compressed digital voice signal containing the supplemental data embedded therein, thereby enabling user decoding to extract the supplemental data while receiving the transmitted voice signal.

From another viewpoint, the invention embraces a method of embedding supplemental data in a digital phone signal that is to be transmitted and received in a system by user voice phone handsets inter-connected in the system through a central station, and without affecting the backwards compatibility of the digital phone signals, that comprises, converting the voice signal, either at the central station or at the user handset, to an intermediate representation thereof by applying an encoding transformation to the voice signal to create resulting floating-point coefficients, but without yet performing quantization and truncation steps that are necessary to convert this coefficient representation ultimately into a compressed discrete digital signal; selecting predetermined portions of the transformed voice signal that are each to contain a bit of the supplemental data; performing quantization and truncation by coefficient domain parity encoding technique that modifies the coefficients so that the resulting quantized and truncated compressed version of the digital signal contains the embedded supplemental data; and transmitting such compressed supplemented signal in the manner of a normal digital phone signal either from the central station to the user handset or from the user handset to the central station, respectively.

Preferred and best mode embodiments, designs and techniques are later presented in detail.

### Drawings

The invention will now be described in connection with the accompanying drawings, Figure 1 of which is a block and flow diagram illustrating an overview of the preferred data encoding process and system of my earlier copending application Serial No. 09/389,941 adapted for use in the cellular phone network system of the present invention;

Figure 2 is a similar diagram presenting an overview of the preferred decoding of the compressed voice signal embedded with the supplemental data of Figure 1, as received by a phone handset user and/or a central station;

Figure 3 is a view similar to Figure 1 showing the use of the previously (and later) discussed steganographic techniques in the encoding process;

Figure 4 illustrates an exemplary signal waveform and Fourier transformation-based compressed coefficient-based representation of the voice signal for use in the coefficient-domain parity encoding process useful with the invention;

Figure 5 is a somewhat more detail block and flow diagram specifically directed to the cellular network application of the present invention, illustrating the single process step of the embedding of data supplemental to the digital voice signal at a point where the signal is converted from an uncompressed representation to a highly compressed digital representation prior to transmission from the user to the all system and from the cell system to the handset user;

Figures 6 and 7 are similar to Figure 5 but are directed to the embedding of data at a central digitizing point and a user's cell phone, respectively;

Figure 8 is a similar diagram applied to the embedding of data in an already compressed signal;

Figure 9 is directed to the extraction of the embedded data from the compressed signal;

Figures 10, 11 and 12 respectively illustrate data embedding using time-domain waveform encoding, frequency-domain waveform coding and Vocoder coding; and

Figure 13 shows an exemplary supplemental screen advertisement displayed at the handset.

#### Description of Preferred Embodiments Of The Invention

As before discussed, the present invention is concerned with data embedding in digital phone signals, as a cellular phone network systems and the like, and without affecting the backwards compatibility of the digital phone signal.

The technique can embed supplemental data into the phone signal, at the user's end for reception at a central station, at the central station for reception at the user's handset, or (with less efficiency) at any intermediate point. This also allows for the creation of a two-way network connection while the handset is used over a voice-only network.

The following types of data are examples of what can be embedded in such a digital phone signal:

From a central station server:

- Individually targeted advertising images which update while user is using the phone.
- Inject market-localized ads or solicitations, as in Figure 13.
- Interactive computer programs such as e-commerce applications, polls, games, or forms.
- Supplemental text or audio content (weather as, for example in Fig. 13; news, pager messages, translations, service updates).
- Music or other entertainment content; call-waiting music and messages, etc.
- Wireless application protocol (WAP) for sending Internet content, two-way.

From a user's handset:

- GPS additional data
- Typed or keyed responses (data backchannel)
- A still image, video or audio channel
- WAP

As shown in Figure 5, the data embedding process consist of a single process added at a point where the audio voice signal is converted from an uncompressed representation to a highly compressed digital representation. This process adds data to the voice signal as part of the coding and compression process, before it is transmitted.

There are two such points in a typical wireless system:

There are two such points in a typical wireless system:

- One such point is where the central cellular phone system receives signals from outside source, typically a public switched telephone network (PSTN). This point is typically the Mobile Switching Center (MSC), in most cellular phone systems.
- The second such point is where the cellular phone converts the user's voice for transmission to the cellular phone system.

At either point, arbitrary data can be placed into the audio stream.

In Figure 5, in the left-hand column, the embedding of supplemental user data from the user to the cell system is shown; transmitting to and receiving by cellular receivers and extracting the supplemental data while reconstructing and presenting the original user voice signal and retransmitting to, for example, PSTN. The right-hand column, from the bottom upward, shows operation from cell system to the user.

Data, as before mentioned and shown at the bottom of Figure 5, can also be added at any point to a previously compressed signal using the same techniques, but typically at a lower bit rate.

As earlier discussed, the supplemental data may be embedded at a central digitizing point. Figure 6 illustrates such a process of embedding data at a central digitizing point for extraction at the user handset. The required method steps are described in sufficiently generic terms to apply to all types of known coders used to compress speech data in cell phones.

The embedding process begins with two components: an audio voice signal and a supplemental data file to be embedded in the audio signal.

The first step, as shown in Figure 6, is to convert the voice signal to an intermediate representation, which depends on the actual coder used. This typically consists of applying the encoding transformation to the voice signal at T, resulting in a set of floating-point coefficients but, without yet performing the ultimate quantization and truncation steps necessary to convert this coefficient representation into a compressed discrete digital signal. Such a Sine, Wavelet or related discrete transform is illustrated in the signal waveform and coefficient-based tabular illustration of Figure 4.

The second step is to choose which portions of the transformed voice signal are each to contain a bit of the supplemental data file. Typically, this is done by selecting at S, a set of coefficients, preferably at regular intervals in the data.

At this point, the previously mentioned coefficient domain parity encoding technique of Figure 4 may be used to modify the coefficients so that the quantized and truncated version of the digital signal at Q, Figure 6, contains the embedded data. The digital data signal may now be transmitted at Tx as a normal digital phone signal.

The thusly compressed voice signal is diagrammatically shown in Figure 1 as combined in an encoding process (so -labeled) of any well-known type, later more fully discussed, with the supplemental data content ("Data") for embedding therein. There then results, a compressed voice signal with supplemental embedded data without affecting its backwards compatibility with existing file formats, and without substantially affecting the handset phone user's receiving or playback experience. If desired, moreover, the transformation step of Figure 1 may be made part of the encoding process, and may even include an optional compression step; or these may be applied as additional separate



steps. In the event that such transformation, compression and encoding processes are combined, indeed, it is then possible to use perceptual encoding techniques to choose into which coefficients to embed the data.

Continuing with the broad overview, however, the decoding and playback are diagrammed in Figure 2, wherein the decoding process, so-labeled and later more fully discussed, is dependent upon the type of encoding process used in Figure 1 to embed the supplemental data. Typically, such involves a simple reversal of the encoding process, as is well-known. The voice signal, as shown, is left unchanged in the decoding process. If desired, moreover, the supplemental data, may be verified (" Verification Process") by well-known checksum or digital signature to insure that the data is bit-wise identical to the data which was originally encoded and embedded in Figure 1.

In the voice signal receiving environment, moreover, the receiving handset or station and the execution environment may communicate with one another, illustrated schematically in Figure 2 by the SYNC line between the voice handset or station receiver and the data manipulation environment boxes, so that the execution of the supplemental data can be synchronized with the reception content.

The possible use of data encoding using steganographic techniques was earlier mentioned with reference citations, and such an application to the techniques of the present invention is illustrated in Figure 3. The supplemental data to be embedded is there shown transformed into a bit stream code, with the bytes of the data extracted into a bit-by-bit representation so that they can be inserted as small changes into the voice signal. The selection of the appropriate locations in the voice signal content into which to embed the data bits, is based on the identification of minor changes that can be made to

the actual media content with minimal effects to the user's voice signal receiving experience. Such changes, however, must be such that they can easily be detected by an automated decoding process, and the information recovered.

At the step of "Insertion of Executable Code" in Figure 3, any one of a number of steganographic encoding processes (including those of the earlier cited references) may be used. In accordance with the present invention, where the voice signal content is represented as a set of function coefficients, the data bits are preferably embedded by the technique of modifying the least-significant bit of some selected coefficients, as hereinafter also more fully discussed.

The resulting voice signal with embedded executable code is again backwards compatible, with, in some cases, slightly diminished, but entirely acceptable, possible user receiving experience due to the embedding process. In accordance with this invention, more than 3000 bits of data per second has been readily embedded in an audio file encoded at a bit-rate of 128,000 bits/sec.

It is now in order to expand upon the selection of the sets of suitable coefficients of the voice signal transform, preferably at regular intervals, for implementing the data bit embedding in accordance with the present invention. As earlier pointed out, the invention need change only a single bit in a selected coefficient, as distinguished from prior art large - scale ordering changes in the relationships of the coefficients (for example, as in the previously cited Zhao and Koch references). This set can be selected by simply choosing a consecutive series of coefficients in the voice signal. A preferred technique is to choose a set of coefficients which encoded a wide range of frequencies in

the voice signal as discussed in connection with the coefficient-domain parity encoding representation of earlier discussed Figure 4.

For each bit in the data bit stream, the selected coefficient and the next data bit to be encoded are combined, re-scaling the coefficients to encode the bit ("Rescale" in Figure 6). If possible, this is preferably done in conjunction with the quantizing and re-scaling step so that the choice of the coefficient to be modified can be based on the closeness of the original coefficient to the desired value. After such quantizing and re-scaling, furthermore, there is not as much data on which to base this decision.

The re-scaling, moreover, can be done in-place in an already-encoded audio file, with the added constraint of keeping the file size constant. In such a case, where it is not possible to encode the bit just by re-scaling a single coefficient while maintaining the frame rate, multiple coefficients may be changed so that their compressed representation remains of the same length and the audio file is accordingly minimally disturbed. This encoding may be accomplished through an LSB encoding process, or preferably through the LSB parity encoding of Figure 4. Such parity encoding allows more choice regarding the coefficients to be modified.

Referring to the illustrative coefficient-based representation of the table in Figure 4, the parity of the coefficient can be computed by adding them together:

$$12 + 15 + 5 + 3 + 10 + 6 + 12 + 1 = 64.$$

Since 64 is even, the bit value currently encoded in these co-efficients is 0. If, however, it is desired to encode a 1 in this set of coefficients, it is only necessary to make the parity odd. This can be done by choosing any amplitude or phase value, and either adding or subtracting 1. This choice of value can be done arbitrarily, or can be made based on the

types of psycho-acoustic models currently used in the before-discussed MPEG encoding process.

This illustrates the use of parity of the low bits of a series of coefficients in the encoding of the data by magnitude frequency-domain low-bit coding. As an example, assume it is desired to encode a single bit of data information in a series of, say, eight coefficients. In accordance with the invention, instead of simply modifying the low bit of the first coefficient, encoding is affected by modifying the parity of the eight low bits together. The algorithm examines a set of consecutive coefficients, extracts the low bits, and counts how many of them are set. Thus, with the technique of the invention, a single bit of data can be encoded whether the number of set bits is even or odd (the parity). This provides the advantage of providing algorithmic choice in determining which coefficient to modify, if any.

Alternatively, this technique may be applied to a wider range of values, while using higher-order parity. As an example, the same amount of data can be encoded over 32 coefficients as can be encoded over 28-coefficient regions, by adding up the low bits of those 32 coefficients and then computing the result modulo four (the remainder when dividing by four). This provides more flexibility in choosing which coefficients to modify, though it does not allow as much data to be inserted into the stream.

Specific references for fuller details of the above-explained techniques usable in the encoding and decoding process components of the invention, are:

[ISO 8859-1] ISO/IEC DIS 8859-1.  
8-bit single-byte coded graphic character sets, Part 1: Latin alphabet No. 1. Technical committee/subcommittee: JTC 1/SC 2;

[MIME] Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies", RFC 2045, November 1996.  
<url::ftp://ftp.isi.edu/in-notes/rfc2045.txt>; and

[UNICODE] ISO/IEC 10646-1: 1993.  
Universal Multiple-Octet Coded Character Set (UCS), Part 1: Architecture and Basic Multilingual Plain Text committee/subcommittee: JTC 1/SC 2  
<url: <http://www.unicode.org>>.

While the preferred use of least-significant bits of the magnitude or amplitude coefficients of the transform frequency representation of, the compressed voice signal has been discussed, other techniques may also be employed such as phase frequency-domain low-bit coding wherein the least-significant bits of the phase coefficients (Figure 4) of the transform frequency representation of the voice signal are used to encode the program. The implementation is the same except for the use of the phase coefficients to encode data as opposed to the magnitude coefficients -- and, in the case of audio content, because the human ear is much less sensitive to the phase of sounds than to their loudness, less audible distortion may be encountered in reception and playback.

The present invention has been illustratively described in Figure 6 for the embedding of the supplemental data at a central digitizing point or server station. It has been earlier mentioned, however, that the embedding of the supplemental data may also be effected at the user's cell phone handset. This operation is shown in Figure 7 where the same reference letters have been applied as in Figure 6. Figure 7 illustrates the process of embedding data at a user's handset for extraction at a central point. This is identical to the process of embedding data at a central digitizing point as detailed in connection with Figure 6, except that the embedding process is performed on the handset, and transmits data from handset to the central station.

As for the embedding of data at other points, Figure 8 illustrates how data can be embedded into a digital signal which has already been compressed. Because the encoding process can no longer take advantage of information about the original voice signal, it cannot, however, embed data into the signal with the same transparency and efficiency before-described. It is possible, however, and may often be useful to add data to the signal at this lower bit rate. This consists of examining the digital voice signal and inserting data at regular intervals by modifying the discrete coefficients which represent the voice signal.

Turning again to the data extraction technique useful with the invention, the embedding data can be extracted from the digital voice signal at any point in the process without affecting the digital signal in any way. While previously discussed Figure 2 represents a broad system, Figure 9 is more detailed and specific to the extraction of the supplemental data embedded in the transmitted compressed voice signal of the invention. Where the data is intended to be received by a user's phone, it can be extracted into an appropriate format and displayed, executed, stored, or otherwise handled by the phone, data shown in Figure 9. Where the data is intended to be received and another point in the system, it is extracted into an appropriate format and acted on in the matter depending on the semantics of the systems.

Further details on the Time-Domain Waveform Coding to enable the supplemental data embedding in the compressed voice signal are presented in Figure 10; with more detailed steps in the alternate use of Frequency-Domain Waveform Coding being presented in Figure 11 and of Vocoder coding in Figure 12.

Returning to the encoding process for embedding supplemental data, there are, as before stated, three main classes of coders used to convert data from an uncompressed representation to a highly compressed digital representation for digital phones: Time-Domain Waveform Coders, Frequency-Domain Waveform Coders, and Vocoders.

In the steps of Figure 10, the voice signal is shown subjected to digitization of voice samples (so-labeled), calculation of adjacent sample differences, and selection of a subset of such sample differences. This is combined with the next selected occurring coding from the transform supplemental data bit stream for embedding of such bit using adaptive quantizing. In this example, there results ADPCM compressed voice with the embedded data.

Such Time-Domain waveform coders try to reproduce the time waveform of the speech signal, are source is dependent, and can thus encode a variety of signals. Examples of these type of coders include pulse code modulation (PCM), differential pulse code modulation (DPCM), the adaptive differential pulse code modulation (ADPCM) above mentioned, delta modulation (DM), continuously variable slope delta modulation (CVSDM), and adaptive predictive coding (APC). All time-domain coders consist of a quantized representation of the waveform.

The ADPCM coder of Figure 10 is widely used in such systems as the PACS (Personal Access Communication Systems) third-generation PCS system, the Personal Handyphone System, and in the CT2 and DECT cordless telephone systems, at a bit rate of 32kbps. A representative system of this type is shown in Figure 10. At this bit rate, it samples the audio stream at 8 kHz, and uses 4 bits to represent the adaptive stepsize differences between each successive audio sample. By embedding data in the lowest bit

of these audio samples at a rate of 1 bit per 6 samples, we can embed data at a rate of 1300 bits/sec, which is 10k bytes/minute.

A Frequency-Domain waveform coder divides the signal into a set of frequency components, which are quantized and encoded separately.

The Frequency-Domain coding of Figure 11 is illustrated for a sub-band coded compressed voice with embedded data operation, wherein the digitized voice samples are filtered into sub-bands, and subsets of the sub-band data (at bit rates depending on the particular sub-band) are selected for appropriately embedding the next bit to be encoded in the transformed supplemental data bit stream.

The CD-900 cellular telephone system; for example, uses a type of frequency-domain waveform coder known as sub-band coding. Let us consider a representative sub-band coding system, shown in Figure 11, which has a bit rate of 8.3 kbps. It divides the audio into four sub-bands, the first of which uses 4 bits at 450 samples/sec to encode their frequency range 225-450 Hz, three bits at 900 samples/sec for 450-900 Hz, 2 bits at 1000 samples/sec for 1000-1500 Hz, and 1 bit at 1800 samples/sec for 1800-2700 Hz. Because each frequency range of the signal is encoded separately, we can embed data at different bit rates in each range. By embedding data at 1 bit per 4 samples in the lowest range, then respectively 1 bit per 8 samples, 1 bit per 12 samples, and 1 bit per 16 samples in the highest range, we can embed data at the rate of 420 bits/sec, or 3.1 k bytes/minute.

A similar procedure for a VSELP illustrative example of the use of Vocoding for the supplemental data embedding is presented in Figure 12, wherein the digitized voice samples are analyzed by an RTE-LTP function before selection of a subset of coefficients



based on perceptual importance, before effecting the embedding of the next bit of the transferred data bit stream in quantization of the RTE-LTP coefficients.

Vocoders are based on extensive knowledge about the signal to be coded, typically voice, and are signal-specific. For example, in GSM, the Vector Sum Excited Linear Predictive (VSELP) coder outputs fifty speech frames per second. These speech frames consist of a set of coefficient parameters to the RPE-LTP (regular pulse excited longer-term prediction) function. These coefficients are then quantized and encoded into 260 bits. According to Annex 2 of the Interim European Telecommunication Standard, I-ETS 300 036, "European digital cellular telecommunications system (phase 1): Full-rate speech transcoding," subjective tests have been performed to determine which of these 260 bits are the most perceptually important. The 69 least perceptually important bits are all contained in the "Class II" portion of the bits, which is the last 78 bits of the frame.

The embedding process illustrated in Figure 12 involves embedding data in these 69 bits, at an embedding rate of 1 data bit per 4 coefficients. We can embed 17 additional data bits per frame. This is a rate of 850 bits/sec, in a media stream transmitted at 13kbps. This is equivalent to a 6.2k picture transmitted every minute.

Digital phone signals are subject to interference and fading. Any of a number of common techniques used to reinforce the digital phone signal and to provide for robustness, error detection, and error correction may be used. Such techniques include parity bits, block codes such as Hamming Codes and Reed-Solomon Codes, and convolutional codes. Additionally, retransmission of the data and interleaving of time-delayed versions of the data can improve robustness.

Another technique is used to create a protocol for guaranteed delivery (e.g. based on TCP/IP), or WAP or the like, using the two-way data embedding techniques described previously to establish a *bidirectional data connection*. Such techniques typically reduce the amount of data that can be embedded in a stream, but are essential where digital data and executable programs must be transmitted without error.

Further modifications will also occur to those skilled in this art, and such are considered to fall within the spirit and scope of the present invention as defined in the appended claims.

What is claimed is:

1. A method of embedding supplemental data in a digital phone signal that is to be transmitted and received in a system by user voice phone handsets inter-connected in the system through a central station, and without affecting the backwards compatibility of the digital phone signals, that comprises, converting the voice signal, either at the central station or at the user handset, to an intermediate representation thereof by applying an encoding transformation to the voice signal to create resulting floating-point coefficients, but without yet performing quantization and truncation steps that are necessary to convert this coefficient representation ultimately into a compressed discrete digital signal; selecting predetermined portions of the transformed voice signal that are each to contain a bit of the supplemental data; performing quantization and truncation by coefficient domain parity encoding technique that modifies the coefficients so that the resulting quantized and truncated compressed version of the digital signal contains the embedded supplemental data; and transmitting such compressed supplemented signal in the manner of a normal digital phone signal either from the central station to the user handset or from the user handset to the central station, respectively.
2. The method of claim 1 wherein embedded supplemental data is extracted from the transmitted signal at a desired point of the system receiving the transmitted signal, without affecting the voice digital signal reception.
3. The method of claim 2 wherein the desired point of reception is a user handset, and the supplemental data is extracted at the handset in a predetermined format, and

displayed, executed, stored or otherwise handled at the handset without affecting the voice signal communication thereat.

4. The method of claim 2 wherein the supplemental data is embedded at a central station server and is selected from one or more of individually targeted advertising images, updateable while the user is using the phone; market-localized ads; interactive computer programs such as e-commerce applications, polls, games or forms; supplemental text or audio content such as weather, news, pager messages, translations and service updates; music and other entertainment content; call-waiting music and messages; and wireless application protocols for two-way sending of internet content.
5. The method of claim 2 wherein the supplemental data is embedded at a user's handset and is selected from one or more of typed or keyed responses at a data back channel; GPS positional data; still images, video or audio channel; and wireless application protocol.
6. The method of claim 2 wherein the system is a wireless cellular phone system and the point of encoding for the supplemental data-embedding is where the central station receives voice signals from a switched telephone network such as a cellular phone system mobile switching center.
7. The method of claim 2 wherein the system is a wireless cellular phone system and the point of encoding for the supplemental data-embedding is where the cellular phone converts the user's voice for transmission to the cellular phone system.

8. The method of claim 2 wherein said encoding transformation is effected by one of frequency-domain waveform encoding, time-domain waveform encoding, and vocoding.
9. The method of claim 2 wherein the coefficients are prepared by discrete transforms selected from the group consisting of Fourier, Cosine, Sine and Wavelet transforms.
10. The method of claim 2 wherein the embedding step uses the least-significant bit of the selected coefficients.
11. The method of claim 10 wherein the selected coefficients are chosen at regular intervals.
12. The method of claim 10 wherein said coefficients are selected as one of or both frequency and phase coefficients.
13. The method of claim 10 wherein single bits of data are embedded by computing the parity of the least-significant bits of a group of said coefficients.
14. The method of claim 13 wherein a perceptual encoding technique is used to select which of a group of said coefficients is to be modified by data embedding.
15. The method of claim 14 wherein said parity of the least-significant bits of said group of coefficients embeds a bit of data while minimizing the effect on said user's perception of the voice signal reception.
16. The method of claim 2 wherein steganographic encoding is employed in which the data is transformed into a bit stream, and said portions are selected where the insertion and embedding of supplemental data bits produce minimal effects in the perception of the user during voice signal reception.

17. The method of claim 16 wherein said insertion and embedding is effected at the least-significant bit of selected coefficients.
18. The method of claim 2 wherein steganographic encoding is employed in which the data is transformed into a bit stream; sets of coefficients are selected to encompass a range of frequencies in the voice signal information, and, for each bit in the bit stream, the selected coefficients and the next bit to be encoded are combined to re-scale the coefficients and encode such bit as embedded.
19. A method of embedding supplemental data in a digital phone signal that is to be transmitted and received in a system by user voice phone handsets inter-connected in the system through a central station, and without substantially affecting the backwards compatibility of the digital phone signal, that comprises, converting the voice signal, either at the central station or at the user handset, to an intermediate representation thereof by applying an encoding transformation to the voice signal to create resulting floating-point coefficients; performing quantization and truncation steps necessary to convert this coefficient representation ultimately into a compressed discrete digital signal; selecting predetermined portion(s) of the transformed and compressed voice signal that is to contain a bit of the supplemental data; modifying the coefficients so that the compressed version of the digital signals contains also the embedded supplemental data; and transmitting such compressed-supplemental signal in the manner of a normal digital phone signal either from the central station to the user handset or from the user handset to the central station, respectively.

20. The method of claim 19 wherein the embedded supplemental data is extracted from the transmitted signal at a desired point of the system receiving the transmitted signal, without affecting the voice digital signal reception.
21. A method of embedding supplemental digital data in a digital voice phone signal without affecting backwards compatibility of the digital phone signal, that comprises, transforming the digital voice phone signal into encoded sets of frequency-domain or other transform coefficient representations of said signal; selecting predetermined coefficient portions that are each to contain a bit of the supplemental data; and embedding the bits at the selected portions while compressing the signal to transmit a compressed digital voice signal containing the supplemental data embedded therein for enabling user decoding to extract the supplemental data while receiving the transmitted voice signal.
22. A system for embedding supplemental data in a digital phone signals that is to be transmitted and received in a network by user voice-phone handsets inter-connected through a central station, and without affecting the backwards compatibility of the digital phone signal, the system having, in combination, encoding means for converting the voice signal, either at the central station or at the user handset, to an intermediate representation thereof by applying an encoding transformation to the voice signal to create resulting floating-point coefficients, but without yet performing quantization and truncation that are necessary to convert this coefficient representation ultimately into a compressed discrete digital signal; means for selecting predetermined portions of the transformed voice signal that are each to contain a bit of the supplemental data; further encoding means for performing

quantization and truncation by coefficient domain parity encoding technique that modifies the coefficients so that the resulting quantized and truncated compressed version of the digital signal contains also the embedded supplemental data; and means for transmitting such compressed-supplemental signal in the manner of a normal digital phone signal either from the central station to the user handset or from the user handset to the central station, respectively.

23. The system of claim 22 wherein means is provided for extracting the embedded supplemental data at a desired point of the system receiving the transmitted signal, without affecting the voice digital signal reception.
24. The method of claim 23 wherein the desired point of reception is a user handset, and the supplemental data is extracted at that handset in a predetermined format, and displayed executed, stored or otherwise handled at the handset without affecting the voice signal communication thereat.
25. A system for embedding supplemental digital data into a digital voice phone signal having, in combination, encoding means for transforming the voice signal information into sets of frequency-domain or other transform coefficient representations of the voice signal; means for selecting predetermined coefficient portions that are each to contain a bit of the supplemental data; and further encoding and compressing means for embedding a bit of the supplemental digital data at each such selected coefficient portions to produce a supplemental compressed voice signal containing such embedded data for enabling user decoding and separate extraction at desired points of the system of both the voice signal and the embedded supplemental data.



26. The system of claim 25 wherein the voice signal is generated by a user at the user's handset in real time.
27. The system for claim 25 wherein the voice signal is pre-recorded.
28. The system of claim 25 wherein the first-named encoding means prepares said coefficient by one of Fourier, Cosine, Sine and Wavelet transforms.
29. The system of claim 25 wherein, in operation, the further encoding means uses the least-significant bit of the selected coefficients.
30. The system of claim 29 wherein the selected coefficients are chosen at regular intervals.
31. The system of claim 29 wherein said coefficients are selected on one of or both frequency and phase coefficients.
32. The system of claim 29 wherein the further encoding means embeds single bits of data by computing the parity of the least -significant bits of a group of said coefficients.
33. The system of claim 32 wherein a perceptual encoding technique is used to select which of a group of said coefficients is to be modified by data embedding.
34. The system of claim 33 wherein the further encoding means responds to said parity of the least-significant bits if said group of coefficients to embed a bit of data, while minimizing the effect on said user's perception of the voice signal playback.
35. The system of claim 25 wherein said extraction preserves backwards compatibility of the phone
36. The system of claim 23 wherein the supplemental data is embedded at a central station server and is selected from one or more of individually targeted advertising

images, updateable while the user is using the phone; market-localized ads; interactive computer programs such as e-commerce applications, polls, games or forms; supplemental text or audio content such as weather, news, pager messages, translation and service updates; music and other entertainment-content; call-waiting music and messages; and wireless application protocols for two-way sending of internet content.

37. The system of claim 23 wherein the supplemental data is embedded at a user's handset and is selected from one or more of typed or keyed responses at a data back channel; GPS positional data; still image, video or audio channel, and wireless application protocol.
38. The system of claim 22 wherein the system provides a two-way phone network connection while the handset is used over a voice-only network.

1/13

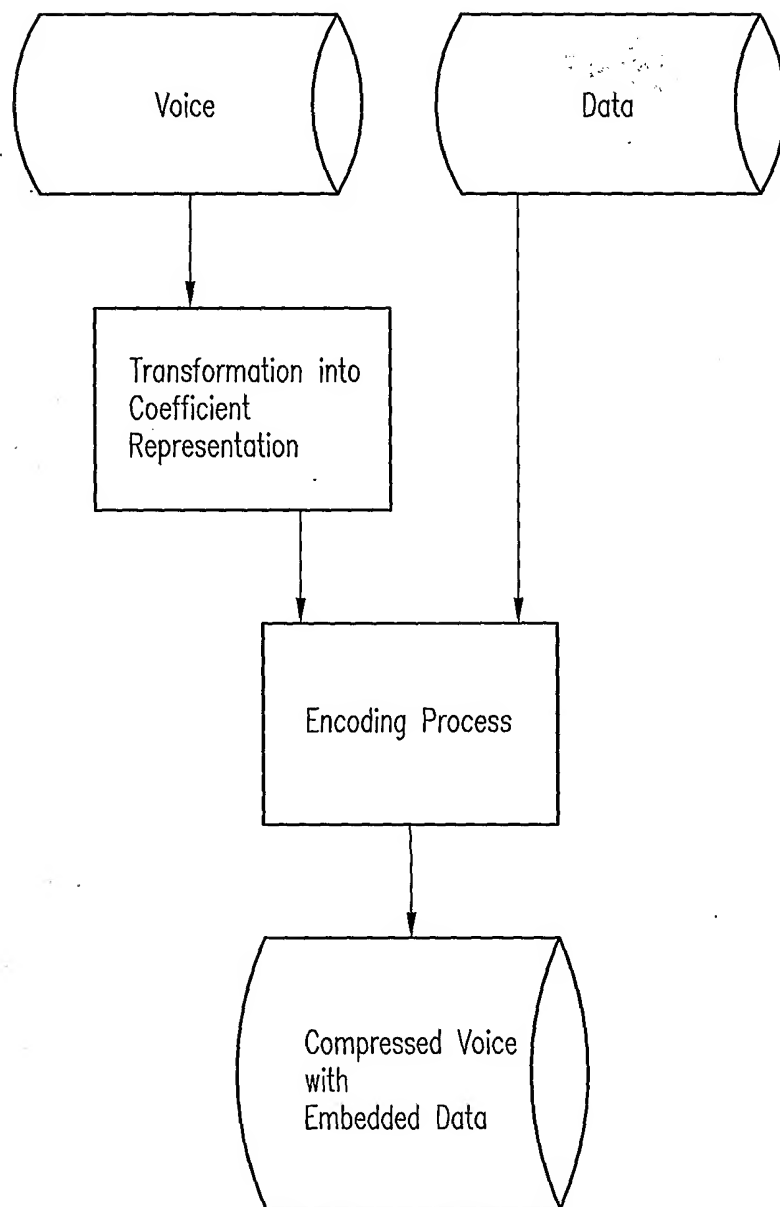
Data Encoding System Overview

FIG. 1

2/13

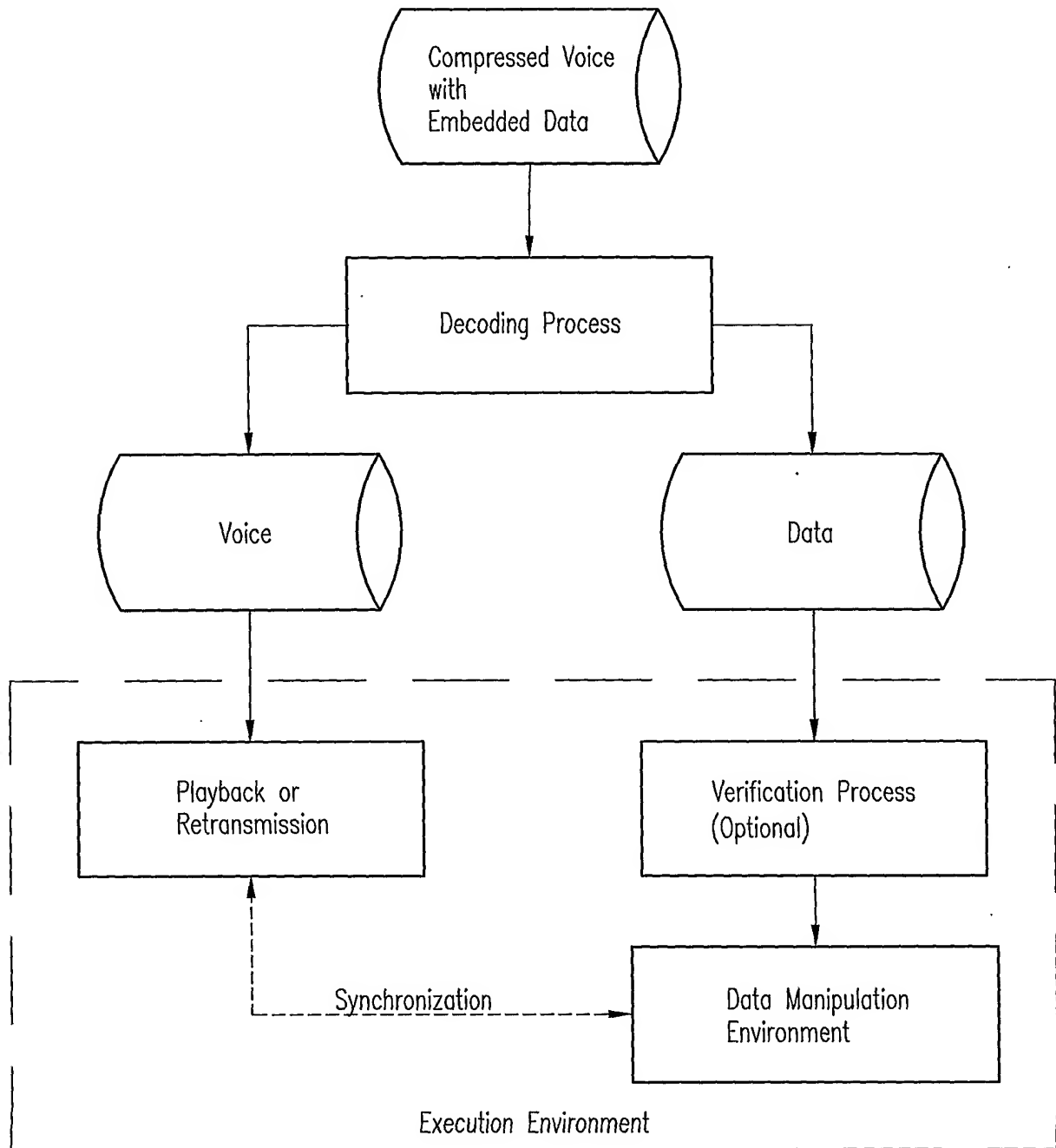
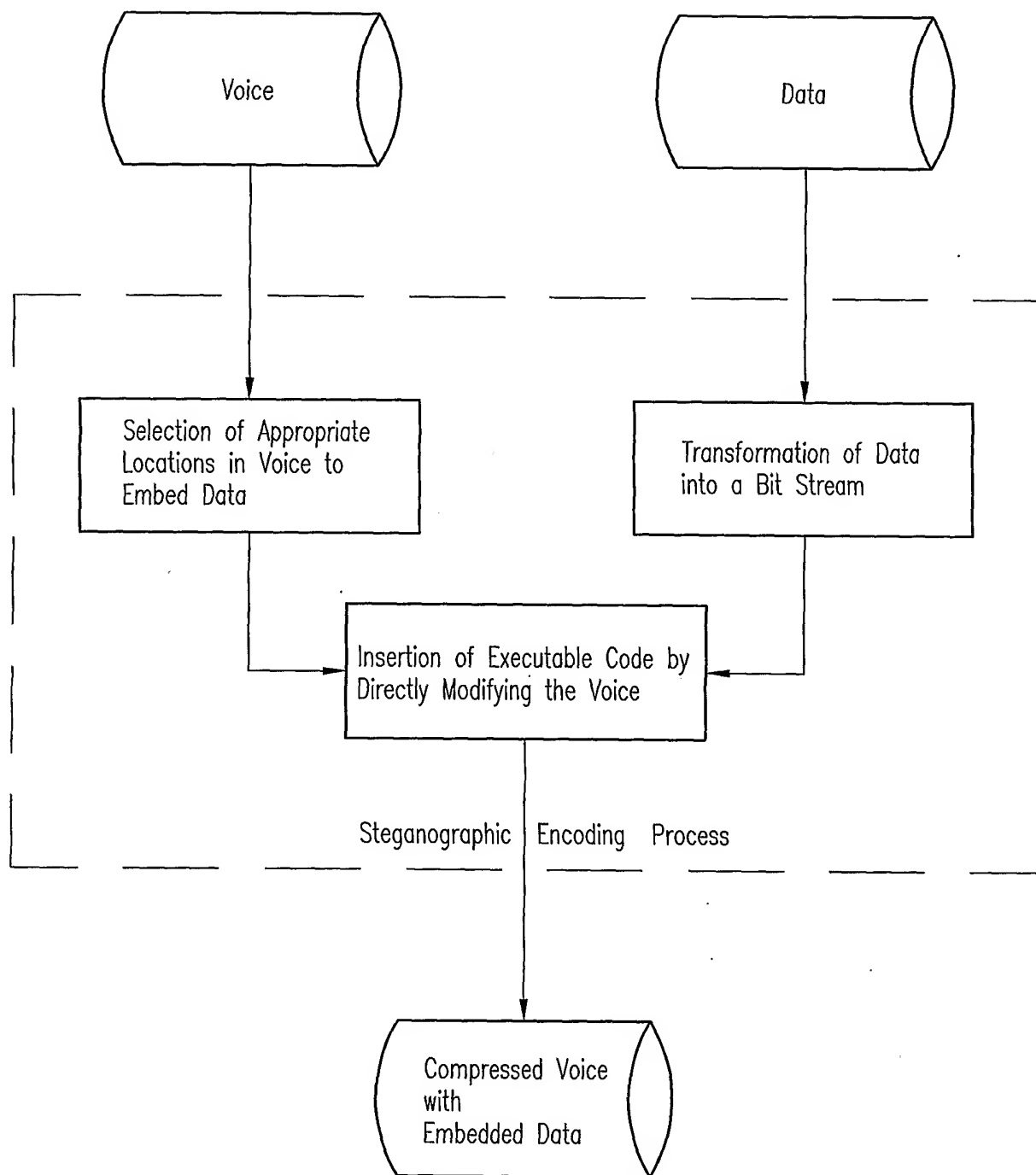
Data Encoding Playback Overview

FIG. 2

3/13

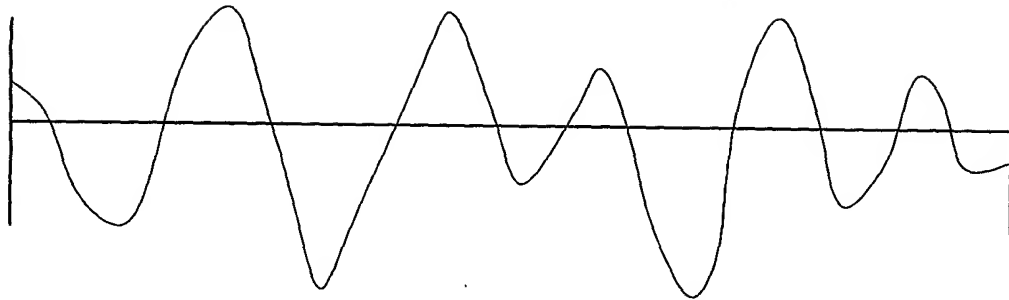


Data Encoding Using Steganographic Techniques

FIG. 3

4/13

Original Signal



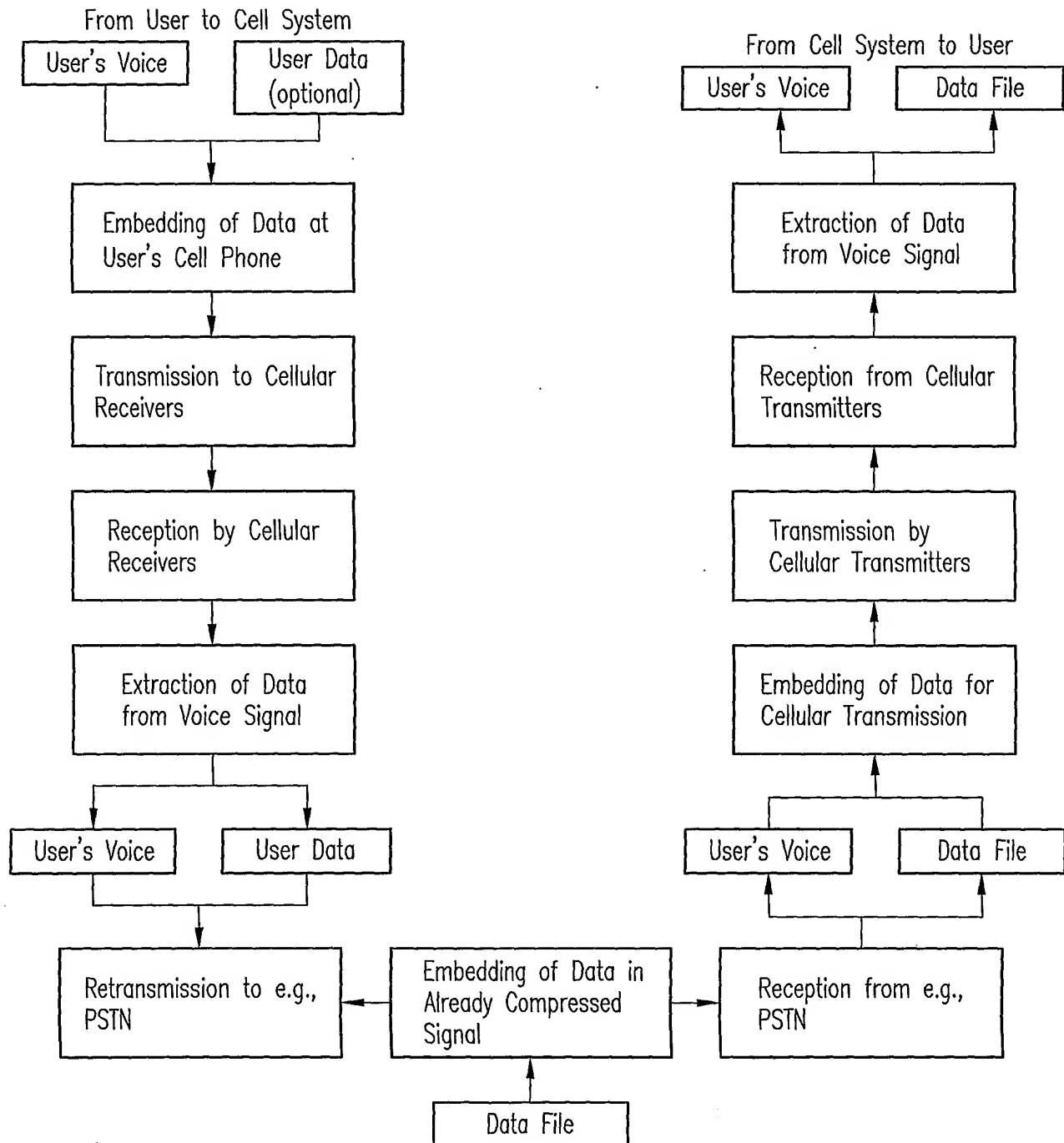
Compressed, Co-efficient Based  
Representation of Signal (In this case  
using a Fourier-Based Representation)

Frequency	Amplitude	Phase
1000	12	10
1200	15	6
1450	5	12
1600	3	1

Coefficient-Domain Parity Encoding

FIG. 4

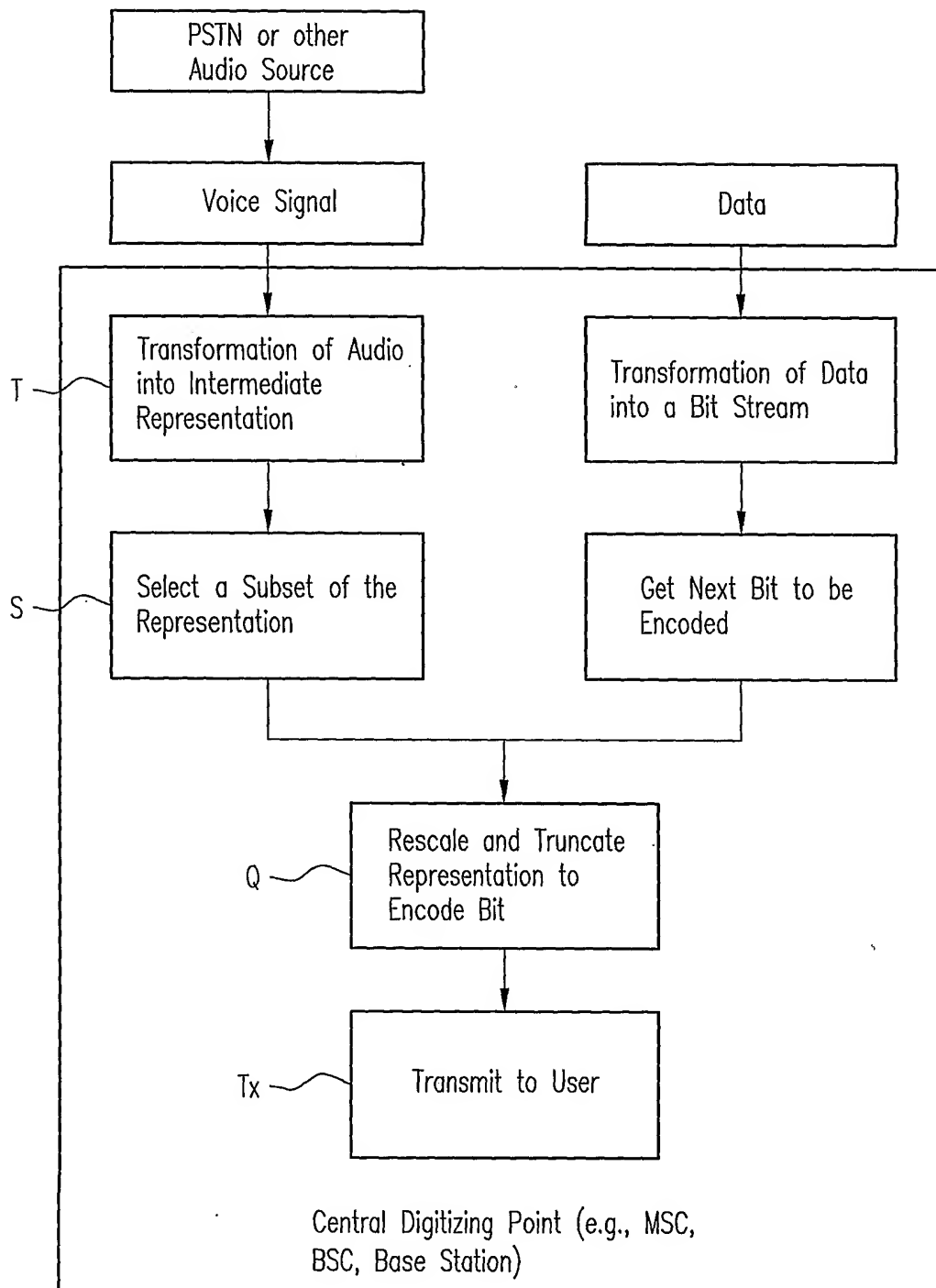
5/13



Cellular Network with Data Embedding

FIG. 5

6/13

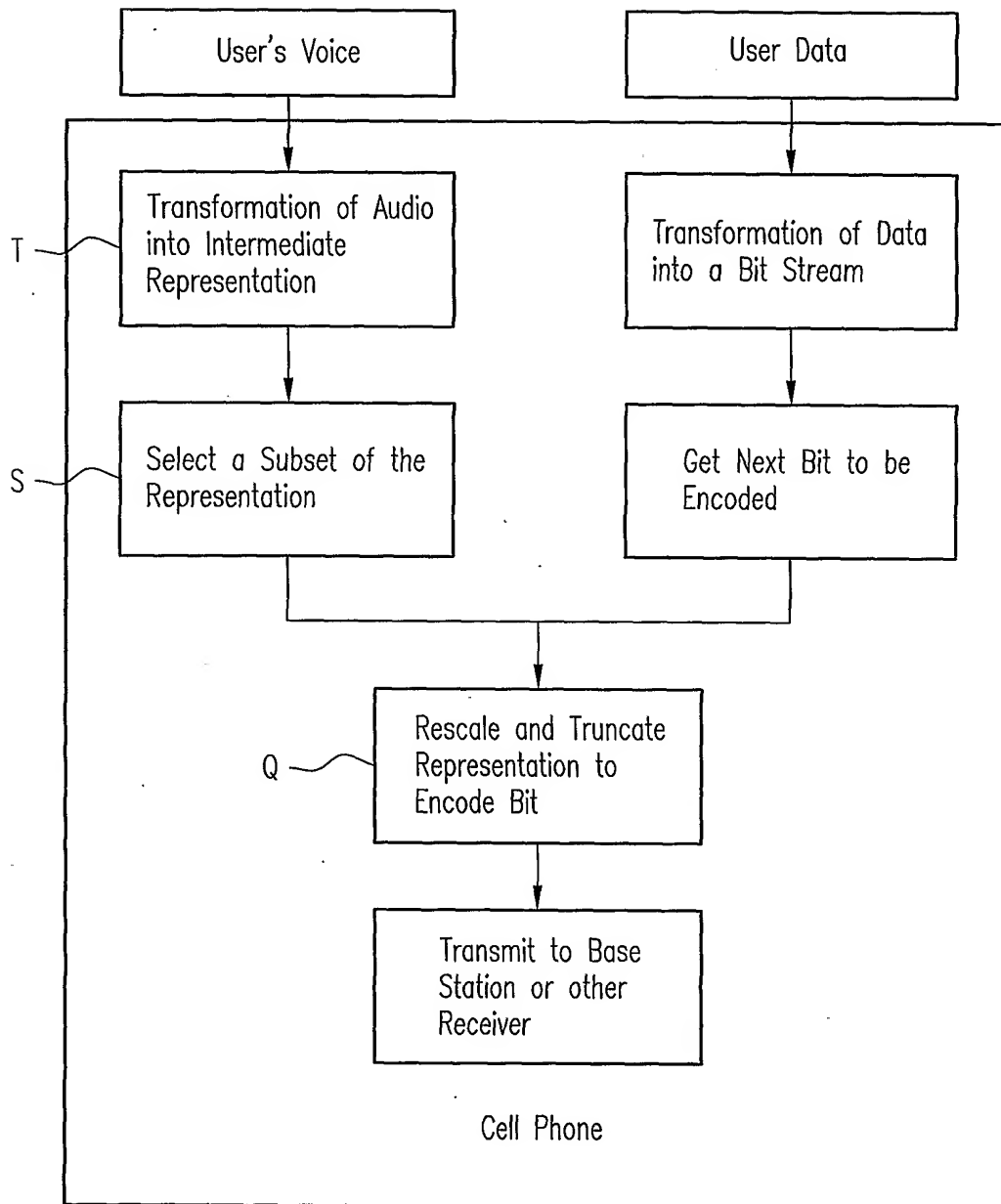


Embedding of Data at Central Digitizing Point

FIG. 6



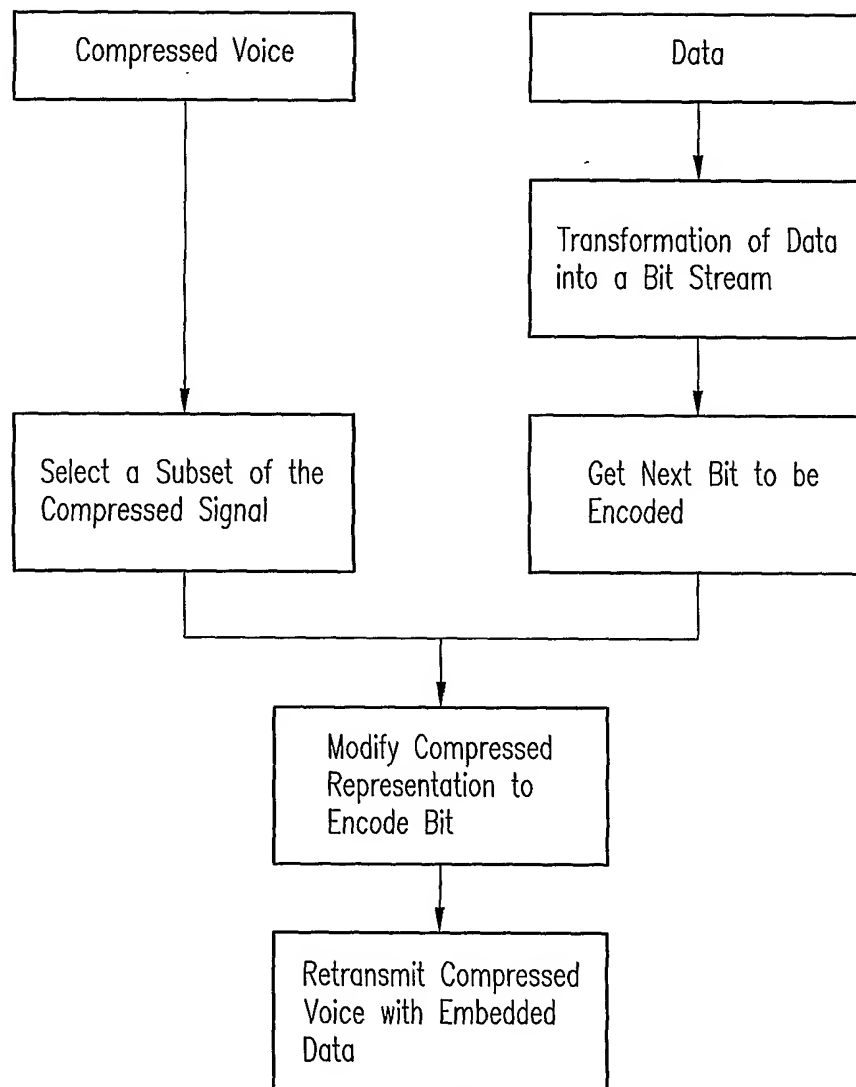
7/13



Embedding of Data at User's Cell Phone

FIG. 7

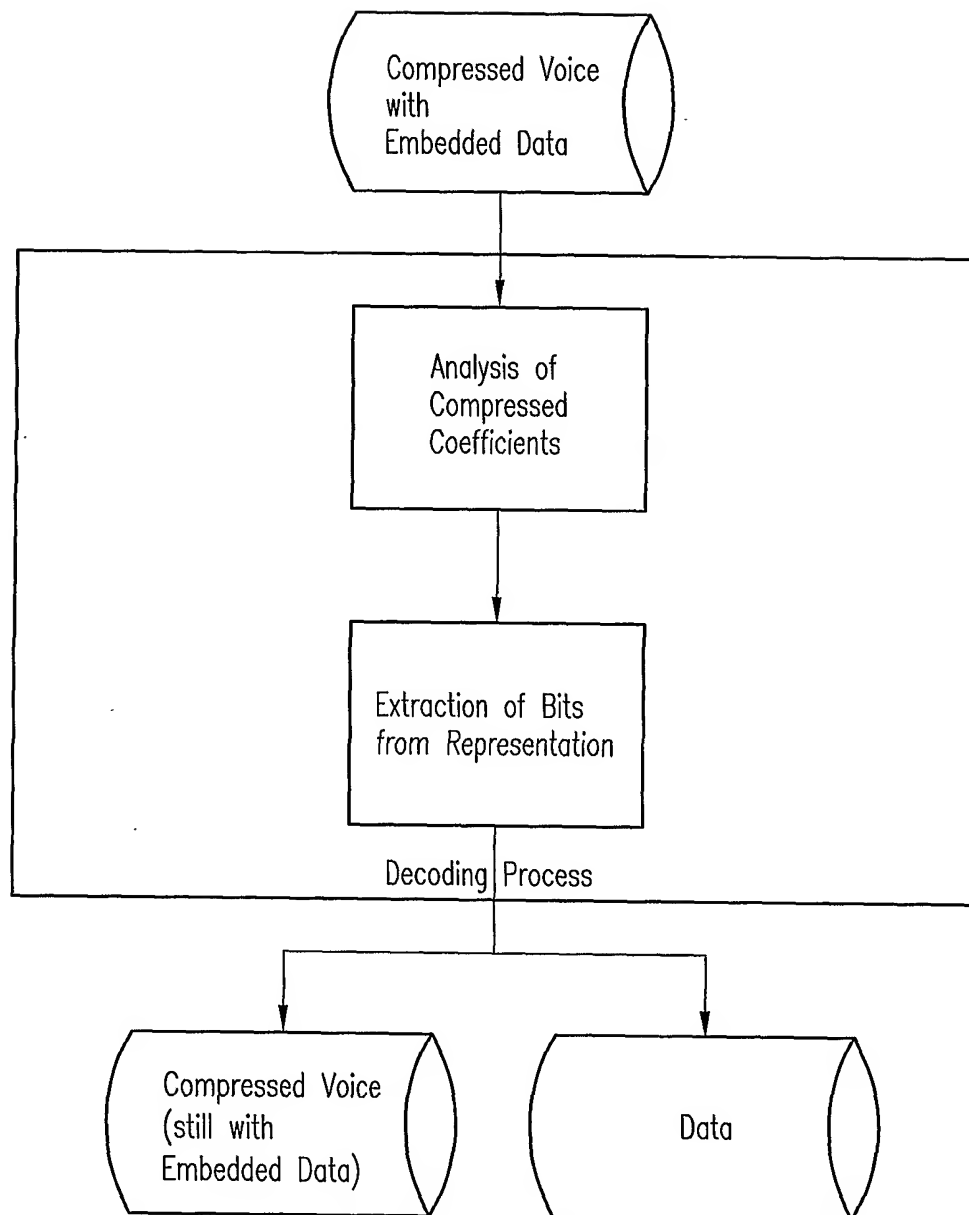
8/13



Embedding of Data in an Already Compressed Signal

FIG. 8

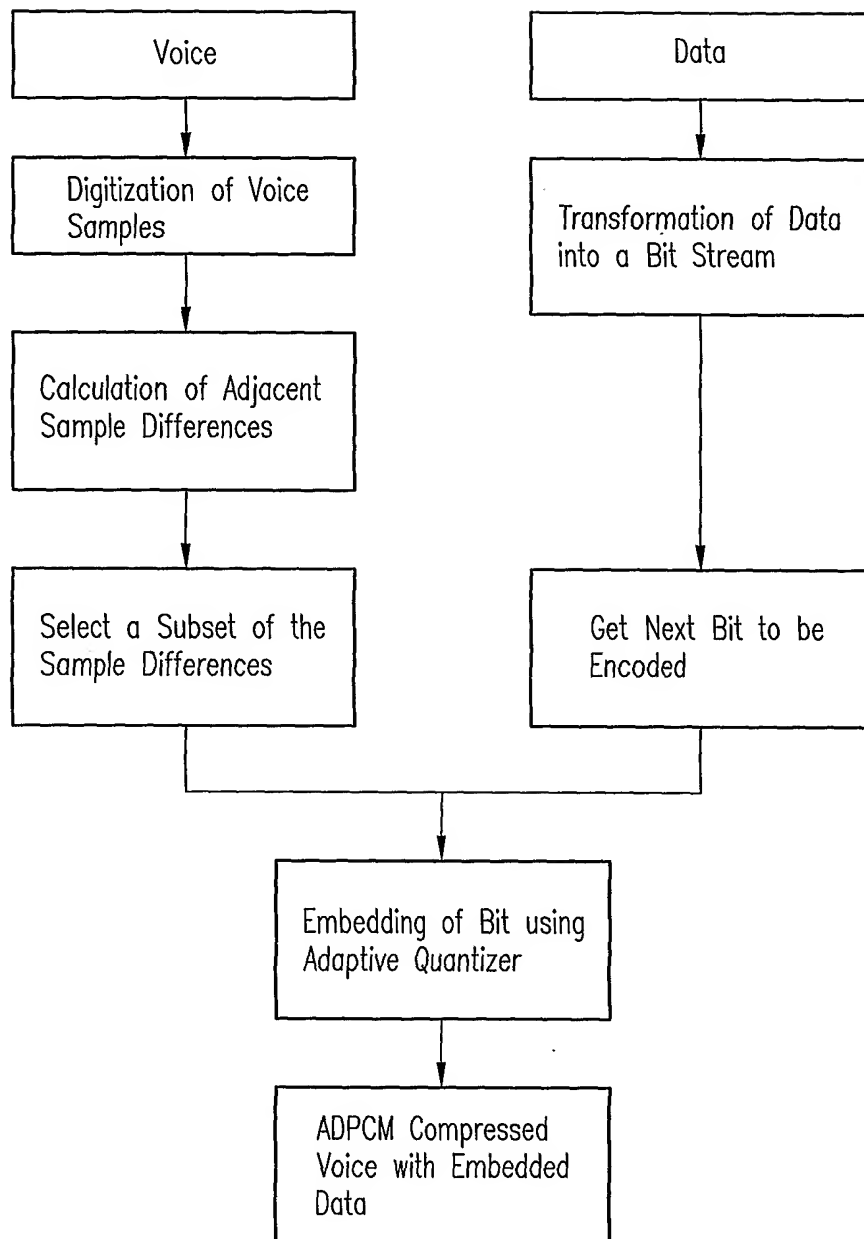
9/13



Extraction of Data from Compressed Signal

FIG. 9

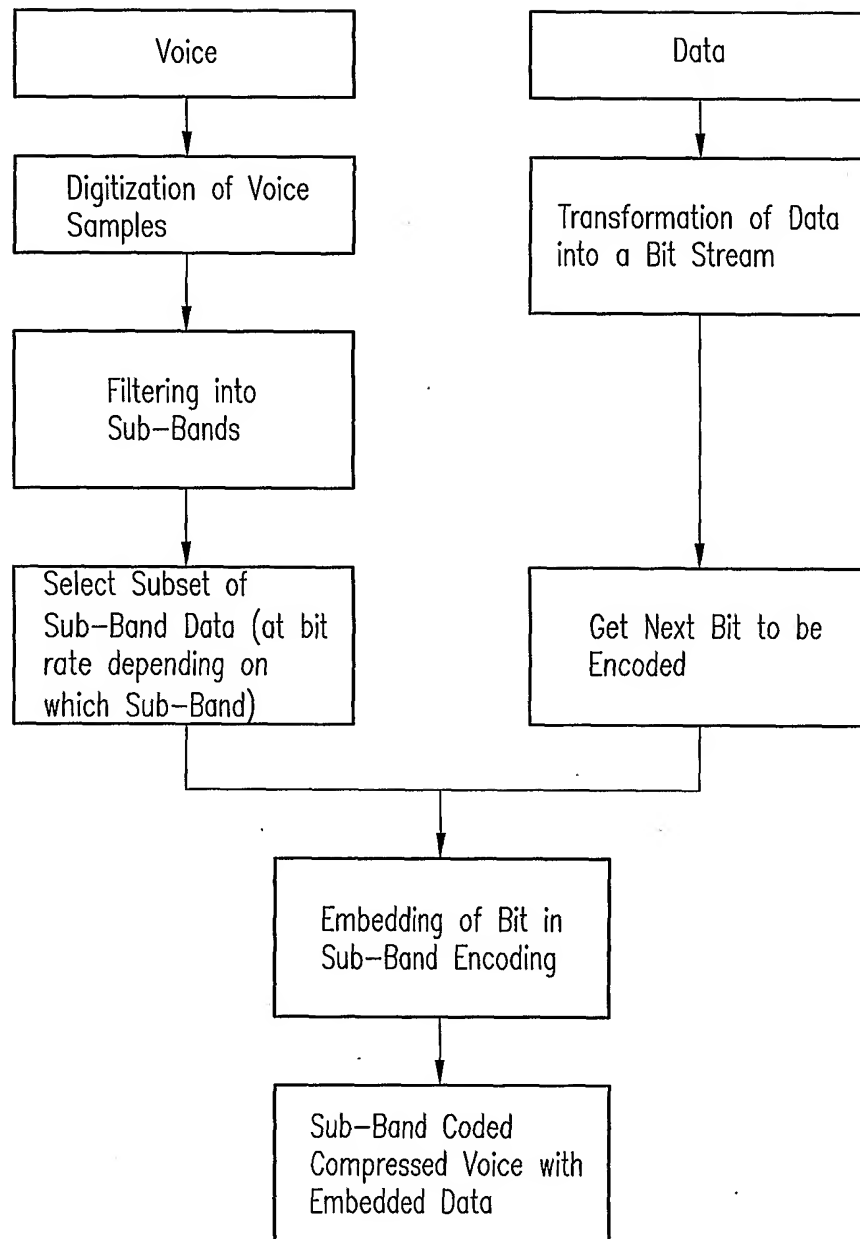
10/13



Embedding Using Time-Domain Waveform Coder  
(ADPCM example)

FIG. 10

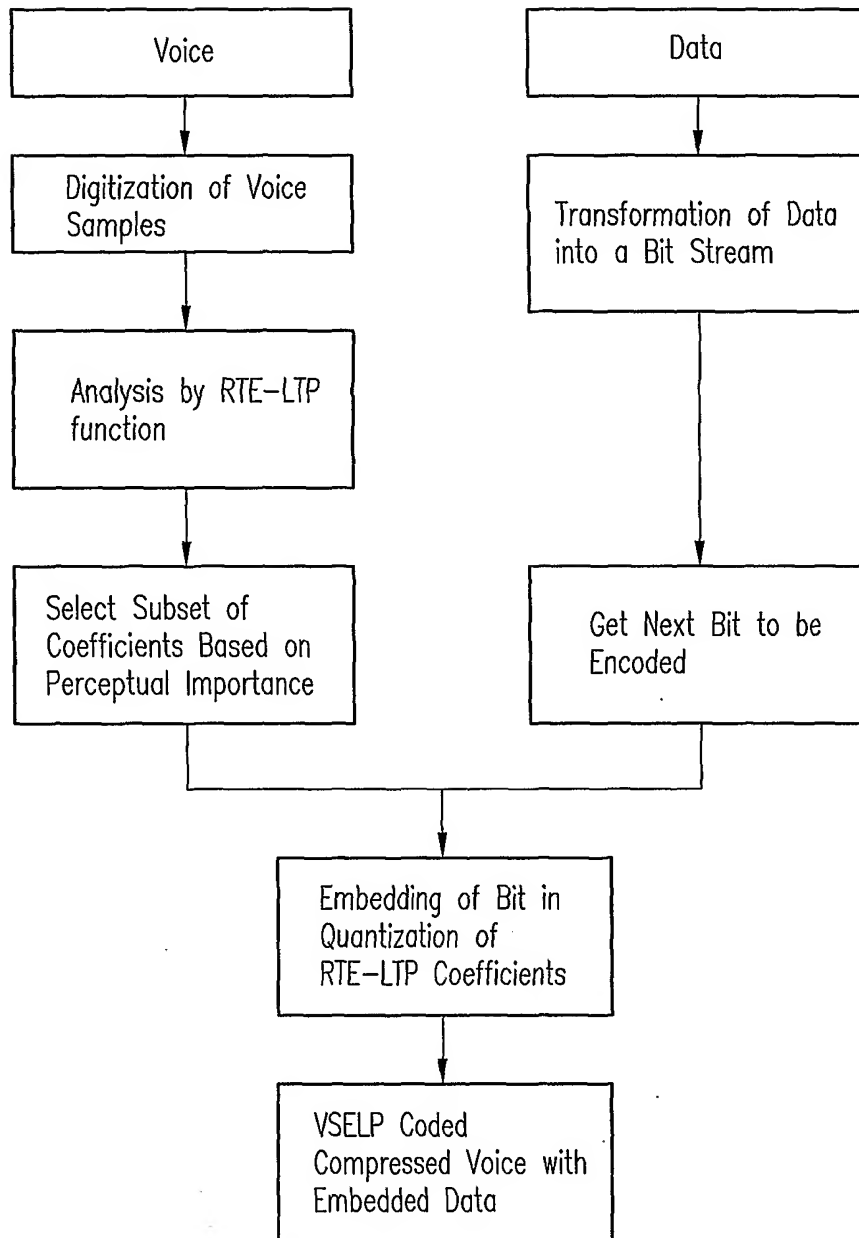
11/13



Embedding Using Frequency-Domain Waveform  
Coder (sub-band example)

FIG. 11

12/13



Embedding Using Vocoder (VSELP example)

FIG. 12

13/13

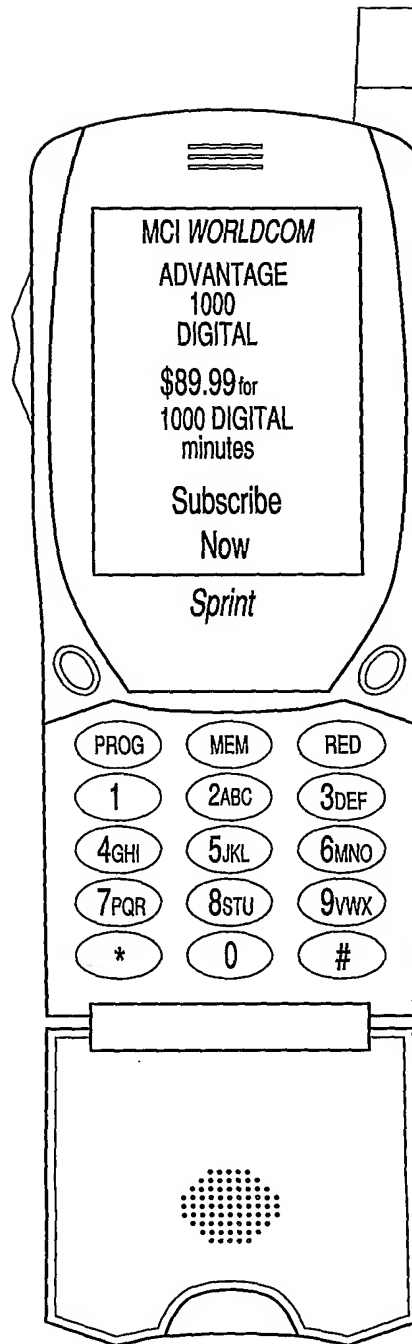


FIG. 13